

Unbiased Group-wise Alignment by Iterative Central Tendency Estimation

M.S. De Craene^{a1}, B. Macq^b, F. Marques^c, P. Salembier^c, and S.K. Warfield^d

^a Center for Computational Imaging & Simulation Technologies in Biomedicine (CISTIB), Networking Biomedical Research Center on Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Information & Communications Technologies Department, Universitat Pompeu Fabra, Barcelona, Spain.

^b Communications and Remote Sensing Laboratory, Université catholique de Louvain, Belgium.

^c Image and Video Processing Group, Technical University of Catalonia, Barcelona, Spain.

^d Computational Radiology Laboratory, Harvard Medical School,
Departments of Radiology, Children's Hospital, Boston, USA.

Abstract. This paper introduces a new approach for the joint alignment of a large collection of segmented images into the same system of coordinates while estimating at the same time an optimal common coordinate system. The atlas resulting from our group-wise alignment algorithm is obtained as the hidden variable of an Expectation-Maximization (EM) estimation. This is achieved by identifying the most consistent label across the collection of images at each voxel in the common frame of coordinates.

In an iterative process, each subject is iteratively aligned with the current probabilistic atlas until convergence of the estimated atlas is reached. Two different transformation models are successively applied in the alignment process: an affine transformation model and a dense non-rigid deformation field. The metric for both transformation models is the mutual information that is computed between the probabilistic atlas and each subject. This metric is optimized in the affine alignment step using a gradient based stochastic optimization (SPSA) and with a variational approach to estimate the non-rigid atlas to subject transformations.

A first advantage of our method is that the computational cost increases linearly with the number of subjects in the database. This method is therefore particularly suited for a large number of subjects. Another advantage is that, when computing the common coordinate system, the estimation algorithm identifies weights for each subject on the basis of the typicality of the segmentation. This makes the common coordinate system robust to outliers in the population.

Several experiments are presented in this paper to validate our atlas construction method on a population of 80 brain images segmented into 4 labels (background, white and gray matters

¹Corresponding author. E-mail: mathieu.dekraene@upf.edu

and ventricles). First, the 80 subjects were aligned using affine and dense non-rigid deformation models. The results are visually assessed by examining how the population converges closer to a central tendency when the deformation model allows more degrees of freedom (from affine to dense non-rigid field). Second, the stability of the atlas construction procedure for various sizes of population was investigated by starting from a subset of the total population which was incrementally augmented until the total population of 80 subjects was reached. Third, the consistency of our group-wise reference (hidden variable of the EM algorithm) was also compared to the choice of an arbitrary subject for a subset of 10 subjects. According to William's index, our reference choice performed favorably. Finally, the performance of our algorithm was quantified on a synthetic population of 10 subjects (generated using random B-Spline transformations) using a global overlap measure for each label. We also measured the robustness of this measure to the introduction of noisy subjects in the population.

Key words: group-wise alignment, probabilistic atlas, expectation-maximization.

AMS subject classification: 93B17, 62G07, 65K10.

1. Introduction

Over the last century, the construction of brain atlases has been essential to understanding human brain anatomy. The initial atlases were developed from cadavers. With the emergence of three-dimensional imaging modalities such as Magnetic Resonance Imaging, research has been aimed at developing 3D digital atlases. To date, most of the available 3D atlases are based on single subjects. The work of Talairach [1], however, has produced a 3-dimensional, proportional system for comparing brains of different sizes in which the brain is organized into areas delimited by anatomical plans. Talairach and others have further defined proportional rules that map coordinates extrapolated from the subject to atlas coordinates. The Surgical Planning Laboratory atlas [2] defines a more complex subdivision of the brain volume obtained by a manual labeling. 3D models are then computed from the label maps and used for visualization and teaching purposes.

The main application of brain atlases reported in the literature is knowledge-based automatic segmentation of a new subject. Such segmentation can be based on non-rigid registration of the atlas on this subject [3] or on hybrid schemes that combine registration and segmentation. A first example can be found in Pohl et al [4] where an Expectation Maximization algorithm is presented to combine the registration of an atlas with the segmentation of magnetic resonance images. When several atlases are available, Rohlfing et al [5] proposed to combine the different classifications obtained by aligning each atlas, weighting the contribution of each atlas by an estimate of the performances of this individual classifier.

Although standard atlases provide a standard system of coordinates, they fail to illustrate the variability of anatomical structures that exists among individuals since a single subject alone cannot represent structural variability. Accurate depiction of variability, on the other hand, will help in the

automatic detection of abnormalities in a pathological image. Atlases computed from a collection of subjects may also enable a differentiation of different populations if significant differences exist in their respective anatomy.

But before approaching the problem of modeling possible anatomical variations across a given population, of all subjects situated in the same space of coordinates, the population must be initially aligned. Two main streams can be identified in the various strategies proposed for joint alignment of a population.

The first stream (*reference-free*) captures the best group-wise alignment by optimizing a global cost function requiring the joint optimization of transformation parameters over the entire population. Studholme [6] simultaneously aligns a group of images to a common space by building a joint histogram whose dimension is equal to the number of subjects in the database. Bhatia et al [7] has proposed the selection of one arbitrary database image to act as an intensity reference. A joint histogram is built with pairs of intensities, each pair comprises the voxel intensity in the reference and the intensity at the same location in each subject. The transformation parameters among all subjects are optimized at the same time to maximize the normalized entropy computed from the joint histogram. Even if these authors demonstrate that such algorithms are suitable for group-wise alignment, the curse of dimensionality (in the joint histogram or optimization space dimensions) may induce convergence problems for large databases. Zollei et al [8, 9] have proposed a congealing approach for dealing with the reference-free alignment of large databases. In their scheme, they minimize the sum of per-voxel entropy by considering separately the contribution of each subject to this global cost function. Marsland et al [10] introduced the notion of *Minimal Description Length* (MDL) for building a group-wise model. The problem of group-wise matching is modeled by solving for the optimal message length encoding the whole population. In this framework, the total message length required for encoding the population is computed as the summation of message lengths for encoding the reference, the set of transformations and the set discrepancy images; discrepancy images being computed as the difference between the reference and each population member. The reference image is computed as the average of all population members. As an alternative reference computation scheme, the authors also consider median filtering for dealing with incomplete data. Unfortunately, if the transformations are effectively computed in an iterative way for optimizing the MDL criterion, the computation of the reference (median or average) is not the result of an MDL optimization process.

Finally, Van Essen [11] has proposed to extract the cortical surface of each subject and to inflate these surfaces to a standard spherical configuration. A reference is computed in the spherical system of coordinates by averaging contours of selected sulci. Each subject is then mapped to this target using landmark-constrained registration.

The second stream (*reference-based*) of algorithms for joint alignment of a population maps a single reference subject to the population being studied (see for instance [12, 13]). The first atlases computed from a collection subjects were produced by averaging 3D MR brain image volumes, each registered to the Talairach [1] atlas. The initial template, referred to as MNI-305 [14], has evolved to the ICBM-152 template that uses higher resolution images acquired as part of the ICBM project [15]. A statistical analysis of the transformations can then be used to characterize different populations as performed in [16, 17, 18].

Within these two categories of algorithms, reference-free and reference-based, most of the authors have focused on the challenge of defining a common space of coordinates that approximates that of an average subject. Indeed, the choice of a non-representative reference would introduce a *bias* in the atlas. By bias, we mean the projection of the population into a common space of coordinates which does not reflect a central tendency in the database, creating this way unnecessary complicated registration problems. If the constructed atlas is biased and depicts peculiar anatomical features for the population being studied, there is a risk that when normalizing a sample of individual brains to this space, systematic deformations could arise, which, in turn, could affect statistical or functional studies [19]. Both in the reference-free and reference-based categories, various approaches have been proposed that would project all population subjects to a common space of coordinates that represents best the population average subject.

First, it is possible to generate a mean transformation by summing all atlas-to-subjects transformations. The magnitude of this mean transformation could then be used as an initial measure of the bias introduced by the current reference. In the reference-free category, both Studholme [6] and Bathia [7] have used this measure of the bias. Studholme [6] incorporates into the cost function a term penalizing large atlas to subjects transformations. Bathia et al [7] explicitly enforces at each iteration the sum of atlas to subjects transformations to be zero. In the reference-based category, Guimond et al [20] and Rueckert et al [21] have used a similar approach. In their work, they picked one of the subjects in the population subject as a reference and then non-rigidly align all other subjects to this subject. After alignment, all subjects are intensity averaged. Last, the mean transformation of all atlas-to-subjects transformations is applied to the sum of aligned subjects, thus removing the bias introduced by the choice of the initial reference.

Averaging subject to atlas transformations as performed in all works referenced in the paragraph above must be performed with caution. It is well known that this approximation is only valid in the small transformation setting. Several contributions have been recently proposed to overcome this limitation and perform averages in the diffeomorphic space of transformations, yielding an accurate estimation of the transformation average, even in presence of large deformations. The reader can refer to Arsigny et al [22] and to Twining et al [23] for a detailed and rigorous mathematical formalization of parametrization and statistics of diffeomorphic transformations.

Another approach is to construct at each iteration a reference reflecting the mean anatomy of the population *under the current alignment*. In this scheme, the group-wise alignment algorithm iterates between the computation of the reference and the alignment of all subjects on this reference. Marsland et al [24] iteratively updates the chosen reference by selecting as reference the subject minimizing the sum of distances between itself and the other ones. Joshi et al [18] address the joint estimation of an intensity template and the transformations that relate the mean coordinate system to those of the population studied. They optimize a two terms metric; one term is the sum of pairwise similarity metrics between the intensity template and each subject; the other term regularizes the transformations by using a Sobolev norm of the diffeomorphism via a partial differential operator on the velocity field. The problem is simplified by noticing that for a set of fixed transformations, the intensity template minimizing the cost function is the intensity average of all subject for the set of transformations. In their implementation section, Joshi et al. describe the algorithm for matching each subject on the average template. The image similarity force is similar

to an optical flow force between the template and the subject. The velocity field is obtained by applying the inverse of the differential Navier-Stokes operator to the force function. Lorenzen et al [25] have used a similar framework but represent each subject by the class posterior mass functions at each voxel rather than using intensities. A Kullback-Leibler divergence measure is used for quantifying the similarity between the template and each subject. The problem is simplified by noticing that for a set of fixed transformations, the template that minimizes the cost function is given by the normalized geometric mean of the deformed class posteriors of each subject. This framework is implemented for a set of two subjects. The framework is shown to be suitable for the matching of multi-modal images by considering in the experiments four imaging modalities: MRA, T1-FLASH MR, T1-MPRAGE MR, and T2 MR.

In this paper, we want to address two questions in the context of aligning a population made up of segmented subjects: Specifically, can we define a common space of coordinates that is unbiased (i.e. a common space of coordinates that does not rely on the choice of an arbitrary subject of the population as reference)?

Further, how can we deal robustly with the presence of outliers in the population that could disturb a reference computation scheme based on the arithmetic mean (e.g. Joshi et al. [18]).

Our algorithm addresses the group-wise alignment by utilizing the Expectation-Maximization framework. The resulting algorithm computes iteratively a reference represented as a probability map for each label. Similar to the STAPLE [26] algorithm, the reference can be seen as the undemocratic mean of the population for the current alignment, giving more weight to consistent subjects in the database (i.e. subjects having a strong correlation between their segmentation and that of the other subjects). The reference is therefore more robust to the presence of outliers than a simple average of all subjects. We measure the similarity between the reference and each subject by building a joint probability distribution and computing mutual information from this distribution.

Next, we optimize mutual information by a stochastic search for affine transformations and by a variational formulation for non-rigid registration. This algorithm has been applied to the alignment of 80 brain scans (Section 3.1.). Section 3.2. quantitatively characterizes the stability of the atlas when the size of the population increases by computing a RMS measure between the atlas computed from a restricted number of subjects and the final atlas estimate. We also compare the group-wise reference computed as a result of our algorithm with a classical reference chosen arbitrarily in the population (Section 3.3.). Finally, we validated the performances of the non-rigid registration algorithm on a synthetic population using a global overlap coefficient. The sensitivity of the performances regarding the inclusion of outliers in the population is also investigated (Section 3.4.).

2. Method

2.1. Definitions and Notations

In this section, we introduce the main notations as well as a framework that addresses the alignment of a collection of labeled images into the same coordinate space. By extending the framework

introduced in STAPLE [26], we advance the validation of image segmentation performed by different experts and/or algorithms. As in STAPLE [26], the consistency of each segmented image, when compared to a central tendency, is measured using transitional probabilities. These probabilities represent the conditional probability of observing a label s' in the subject when the label at the same point in the reference is s .

The reference is not an arbitrary subject of the population, but is modeled as the hidden variable of an EM algorithm. The EM algorithm is an effective strategy for identifying the local maxima of an incomplete-data log likelihood function. It proceeds by assuming the existence of unknown hidden variables which together with the observed data form the so-called ‘complete data’. The expected value of the complete data log likelihood function is computed and the values of the parameters maximizing this function are then obtained. Alternating estimation of the expected value of the complete data log-likelihood function based on estimates of the hidden variables, and maximization of the function then proceeds.

Let θ denote the parameters, and seek the value of these parameters that maximizes the function

$$\theta_{opt} = \arg \max_{\theta} \ln f(\mathbf{W}, \hat{\mathbf{D}}|\theta) \quad , \quad (2.1)$$

where the atlas \mathbf{W} models the unknown most typical anatomy of the population of aligned images. $\hat{\mathbf{D}}$ stands for the collection of segmented images to be aligned and has the dimension of $N_{\text{voxels}} \times N_{\text{subjects}}$. In Equation (2.1), f represents the probability density function of the complete data $(\mathbf{W}, \hat{\mathbf{D}}|\theta)$. The vector θ concatenates two set of parameters

- θ_{T^κ} : *transformation parameters*. These parameters characterize the set of transformations T^κ for $\kappa = 1 \dots N_{\text{subjects}}$ bringing each subject into alignment with the atlas coordinate system
- $p^\kappa(s'|s, \theta_{T^\kappa})$: *similarity parameters*. These conditional probabilities measure the degree of similarity between the segmentation of each subject and the atlas [26]

For convenience, we will use superscript Greek letters indices for counting the subjects of the population and Latin subscript for counting voxels. Bold letter (\mathbf{D}) will designate the collection of voxel values over the entire image whereas non-bold letter will indicate realizations of scalar variables at a given voxel i (D_i). Moreover, for emphasizing that all label values are conditioned over the set of current transformations and must be computed by interpolation, we will use the $\hat{\mathbf{D}}$ notation to indicate the application of the interpolation operator on the segmented image \mathbf{D} . The interpolated label at voxel i for subject κ will therefore be written as \hat{D}_i^κ .

2.2. EM Group-wise Registration Framework

In this section, we examine how the EM algorithm in STAPLE can be adapted for performing the joint alignment of collected segmentations of the same tissue classes acquired from different subjects.

As in any EM algorithm, our algorithm iterates between an Expectation step (*E step*) computing the reference for a given set of subjects - using an estimate of transformations and similarity

parameters - and a Maximization step (*M step*) optimizing a modified likelihood function. Alternating between reference estimate and alignment of all subjects is common to other group-wise alignment algorithms [20, 21]. The originality of our approach is to derive the reference and the optimal set of transformations from a common EM framework.

2.2.1. E Step

This step estimates the *a posteriori* density of the hidden variable conditioned over the observations of subject labels and the current estimate of similarity and transformation parameters. It is common to use the Bayes rule in this step to express the posterior probabilities $p(\text{hidden variable} \mid \text{observations})$ as a function of $p(\text{observations} \mid \text{hidden variable})$ that are related to the model used to compute the likelihood cost function.

Since both the transformation parameters and the similarity parameters are held constant in this step, the probability of observing label s at voxel i in the atlas (W_{si}) can be computed as in [26] using

$$W_{si} = \frac{\pi_s}{C_i} \prod_{\kappa}^{N_{\text{subjects}}} p^{\kappa}(\hat{D}_i^{\kappa} | s, \theta_{T^{\kappa}}) \quad , \quad (2.2)$$

where the similarity parameter $p^{\kappa}(\hat{D}_i^{\kappa} | s, \theta_{T^{\kappa}})$ represents the conditional probability of observing a label s' in the subject κ when the label at the same point in the reference is s , π_s is the prior probability of label s in the atlas and C_i is a normalization constant ensuring $\sum_s W_{si} = 1$ at each voxel.

2.2.2. M Step

In this step, the transformation parameters and the similarity parameters are jointly optimized. The cost function is a modified likelihood score given by

$$\theta_{opt} = \arg \max_{\theta} E_W \{ \ln p(\hat{\mathbf{D}} | \mathbf{W}, \theta) \} \quad , \quad (2.3)$$

where the $E_W\{\cdot\}$ denotes the expectation operator taken over the per-voxel label probabilities in the atlas (Equation 2.2).

If all subjects are assumed independent of each other, the global cost function in Equation (2.3) can be reformulated as a sum of contributions from each subject

$$\theta_{opt} = \arg \max_{\theta} \sum_{\kappa} \underbrace{E_W \{ \ln p(\hat{\mathbf{D}}^{\kappa} | \mathbf{W}, \theta_{T^{\kappa}}) \}}_{CF^{\kappa}} \quad , \quad (2.4)$$

where CF^{κ} is defined as the cost function to optimize for each subject κ . Each CF^{κ} term can then be decomposed as

$$\begin{aligned} CF^{\kappa} &= \sum_i E_W \{ \ln p(\hat{D}_i^{\kappa} | W_i, \theta_{T^{\kappa}}) \} && \rightarrow \text{Voxel independence} \\ &= \sum_i \sum_s p(W_i = s) \ln p(\hat{D}_i^{\kappa} | W_i = s, \theta_{T^{\kappa}}) && \rightarrow \text{Def. of expectation operator} \\ &= \sum_s \sum_{s'} \sum_{i: \hat{D}_i^{\kappa} = s'} W_{si} \ln(p^{\kappa}(\hat{D}_i^{\kappa} = s' | s, \theta_{T^{\kappa}})) && \rightarrow \text{Grouping terms.} \end{aligned} \quad (2.5)$$

As previously defined, the similarity parameter $p^\kappa(s'|s, \theta_{T^\kappa})$ stands for the probability of observing a label s' in the subject κ when the atlas label at the same point is s . The difference between bold and non-bold variables, to distinguish between image and voxel contributions, as well as the hat operator, for referring to interpolated values, have been defined at the end of Section 2.1.

Equation 2.5 introduces the conditional probability of observing a label at a voxel i over the distribution of atlas probabilities at this same location. The presence of conditional probability reminds of a popular similarity used for image registration: the correlation ratio, as introduced by Roche et al [27]. The main difference between the correlation ratio and entropy-based similarity metrics, mainly relies in the dispersion measure applied on marginal and conditional (or joint) probability densities. In our case, entropy has the advantage of appearing as the natural consequence of the likelihood measure defined in Equation (2.1), whereas the use of the variance dispersion measure would be difficult to justify on a theoretical point of view.

The CF^κ should be optimized over the transformation parameters θ_{T^κ} (which will modify the label values \hat{D}_i^κ at each voxel i of the atlas) and over the similarity parameters $p^\kappa(s'|s, \theta_{T^\kappa})$ for each subject. The constraint that $\sum_{s'} p^\kappa(s'|s) = 1$ can be taken into account using a Lagrange multiplier.

Previously, it has been shown [26] that for a given alignment, the derivation of each CF^κ leads to the following solution of the similarity parameters for each subject

$$p^\kappa(s'|s, \theta_{T^\kappa}) = \frac{\sum_{i: \hat{D}_i^\kappa = s'} W_{si}}{\sum_i W_{si}} \quad . \quad (2.6)$$

Thus, for any choice of the transformation set T_j , the similarity parameters for each subject given by Equation (2.6) is a closed form solution of Equation (2.3). If we inject this closed form solution in the cost function to optimize for each subject, Equation (2.5) can be rewritten as

$$\begin{aligned} CF^\kappa &= \sum_{ss'} p^\kappa(s'|s, \theta_{T^\kappa}) \cdot \ln(p^\kappa(s'|s, \theta_{T^\kappa})) \cdot \sum_i W_{si} \\ &= \sum_{ss'} p^\kappa(s'|s, \theta_{T^\kappa}) \cdot \pi_s \cdot \ln(p^\kappa(s'|s, \theta_{T^\kappa})) \\ &= H(D^\kappa | W) \quad , \end{aligned} \quad (2.7)$$

where W is the probabilistic atlas and π_s has been defined as $\sum_i W_{si}$. An additional normalization of π_s is required for having π_s representing a prior probability but this only affects the cost function by a multiplicative constant.

Equation (2.7) shows that the cost function optimized in the EM framework can be interpreted as the conditional entropy of the j^{th} subject with the atlas. Indeed, the $p^\kappa(s'|s, \theta_{T^\kappa})$ parameters represent the conditional probability of observing the label s' in subject κ when the atlas label at the same location is s . The second line of Equation (2.7) is thus the definition of the conditional probability of D^κ over the atlas. This could have been intuitively expected since optimizing a likelihood function and an entropy criterion are related concepts (see Zollei [8, 9]).

The M step can thus be summarized as the iterative optimization over the alignment parameters of the cost function in Equation (2.7) with the subject similarity parameters given by Equation (2.6). Since no closed form solution of the cost function in Equation (2.7) is available for

the transformation parameters, an iterative optimization method is proposed. The optimization methods used for affine and dense non-rigid transformations are described in Sections 2.3.2. and 2.3.3.

2.3. Group-wise Registration Algorithm

In this section, we describe the group-wise registration algorithm derived from the EM group-wise registration framework detailed in the previous section. The cost function described in Equation (2.7) is augmented with another term for making it more robust to global transformations (e.g. affine transformations). We have used two distinct optimization approaches for affine transformation models and dense non-rigid deformations: 1) a stochastic search for optimizing the affine model and 2) a variational analysis for computing a dense non-rigid deformation field. The two approaches are chained for a robust estimation of the group-wise alignment parameters of the population. Finally, for accelerating the convergence of the probabilistic atlas W_{si} (Equation (2.2)), the EM iterative scheme has been slightly modified as described in Section 2.3.4.

2.3.1. Stabilizing the cost function for global alignment

To ensure that the cost function can be utilized for both affine and local group-wise alignment, a second term to the conditional entropy must be added. Indeed, by minimizing Equation (2.7), the optimization process may lead to an unintentional trivial solution of little import. The following inequality

$$0 \leq H(\mathbf{D}^\kappa | \mathbf{W}) \leq H(\mathbf{D}^\kappa) \quad (2.8)$$

shows that $H(\mathbf{D}^\kappa | \mathbf{W})$ can be minimized simply by reducing its upper bound $H(\mathbf{D}^\kappa)$. Thus, a set of transformations moving all subjects outside the atlas domain would result in zero entropy for each subject, effectively minimizing the criterion in Equation (2.7) without yielding a satisfactory solution. This unwanted solution can be discarded by adding the marginal entropy of the κ^{th} transformed subject's image to the cost function. The criterion needed to optimize is thus the mutual information between the subject and the probabilistic atlas

$$MI^\kappa = H(\mathbf{D}^\kappa) - H(\mathbf{D}^\kappa | \mathbf{W}) = H(\mathbf{D}^\kappa) + H(\mathbf{W}) - H(\mathbf{D}^\kappa, \mathbf{W}) \quad , \quad (2.9)$$

where $H(\mathbf{D}^\kappa)$ is measured using the marginal histogram of the subject labeled image.

2.3.2. Group-wise Affine Alignment

First, an affine transformation model is used to increase the similarity between the atlas and each subject. The set of transformations is initialized by translations that map the center of gravity of each subject's segmentation to the center of the common space of coordinates. As a critical aspect of optimization, continuous changes in cost function must occur in parallel with continual changes in the transformation parameters.

A simple way of achieving this is to substitute the computation of $p^\kappa(s'|s)$ in Equation (2.6) by

$$p^\kappa(s'|s, \theta_{T^\kappa}) = \frac{\sum_i W_{si} \cdot \beta(\hat{D}_i^\kappa - s')}{\sum_i W_{si}} \quad , \quad (2.10)$$

where β is a continuous kernel function. It can be seen that resetting β to a Kronecker function in Equation (2.10) is equivalent to Equation (2.6). This is actually equivalent to perform a Parzen window estimation of the joint probability density function, similarly to the mutual information estimator proposed by Viola et al [28]. The difference with Parzen windowing is that the kernel function is only evaluated at the discrete binned positions in the joint histogram, which avoids the quadratic complexity regarding the number of samples as in classical Parzen windowing. We opted for a BSpline regularization of order 3. This means that the β kernel covers three bins in the joint histogram. One advantage of using a B-Spline kernel is that the sum of increments to the histogram bins for one sample is still equal to 1. The use of this kernel for the regularization of the joint histogram is classic in mutual information based registration and has been proposed by Mattes et al [29]. It is known to improve the continuity of the cost function when combined with a quasi-Newton optimizer or a gradient ascent scheme as in our case.

In our implementation, we have been using the Simultaneous Perturbation Stochastic Approximation (SPSA) introduced by Spall [30]. The main feature of the SPSA algorithm is to provide a stochastic and fast estimate of the cost function gradient in a finite difference way by perturbing simultaneously the entire set of parameters. The gradient estimate only requires two computations of the cost function gradient, no matter the space of the parametric space. Proof of convergence and comparison with standard finite difference schemes can be found in [30]. This optimization method has recently been investigated for optimizing mutual information in image registration problems (as in [31]).

2.3.3. Group-wise Non-Rigid Alignment

For maximizing Equation (2.9), an iterative variational displacement field is derived similarly to the method described by Hermosillo et al [32] and Rougon et al [33]. These two works describe a framework for optimizing an image similarity functional \mathcal{S} depending on a displacement field \mathbf{u} . The displacement field \mathbf{u} is related to the transformation T by $\mathbf{u} := T - Id$ where $T : D \rightarrow D$ belongs to some function space \mathcal{T} and is defined over the image domain $D \subset \mathbb{R}^n$. The first variation $\partial_{\mathbf{u}}$ of $\mathcal{S}(\mathbf{u})$ is obtained by computing the derivative

$$\frac{\partial \mathcal{S}(\mathbf{u} + \epsilon \mathbf{h})}{\partial \epsilon} \quad (2.11)$$

of the perturbed criterion with $\mathbf{h} \in \mathcal{T}$ at $\epsilon = 0$.

In our case, the functional \mathcal{S} to optimize is the cost function MI^κ measuring the alignment between the atlas estimate and the subject κ , as defined in Equation (2.9). Since $H(\mathbf{W})$ is not depending on the transformation parameters, optimizing MI^κ is equivalent to optimizing

$$\mathcal{S} = H(\mathbf{D}^\kappa) - H(\mathbf{D}^\kappa, \mathbf{W}) \quad . \quad (2.12)$$

The computation of \mathcal{S} therefore requires the computation of two probability densities. The first, is the joint probability distribution between atlas and subject labels and can be computed from the similarity parameters $p^\kappa(s'|s, \boldsymbol{\theta}_{T^\kappa})$ by applying the Bayes rule to Equations (2.10) and (2.6) as follows

$$\begin{aligned} p^\kappa(s', s|\boldsymbol{\theta}_{T^\kappa}) &= p^\kappa(s'|s, \boldsymbol{\theta}_{T^\kappa}) \cdot p(s) \\ &= \frac{\sum_i W_{si} \cdot \beta(\hat{D}_i^\kappa - s')}{\sum_i W_{si}} \cdot \frac{\sum_i W_{si}}{\sum_{si} W_{si}} = \frac{\sum_i W_{si} \cdot \beta(\hat{D}_i^\kappa - s')}{\sum_{si} W_{si}} . \end{aligned} \quad (2.13)$$

In the former equation, the prior probability of each label $p(s)$ is estimated from the atlas by averaging the probability of label s over all voxels and normalizing the estimate obtained for each label to ensure that $\sum_s p(s) = 1$. The second probability density required for computing \mathcal{S} is the label probability density $p^\kappa(s'|\boldsymbol{\theta}_{T^\kappa})$ for each subject κ and can be computed by integration of $p^\kappa(s', s|\boldsymbol{\theta}_{T^\kappa})$ over the atlas labels s . Using these definitions of $p^\kappa(s', s|\boldsymbol{\theta}_{T^\kappa})$ and $p^\kappa(s'|\boldsymbol{\theta}_{T^\kappa})$, the functional \mathcal{S} can be rewritten as

$$\mathcal{S}(\mathbf{u}^\kappa) = \sum_{s, s'} p^\kappa(s'|s, \mathbf{u}^\kappa) \cdot \ln(p^\kappa(s'|s, \mathbf{u}^\kappa)) - \sum_s p^\kappa(s|\mathbf{u}^\kappa) \cdot \ln(p^\kappa(s|\mathbf{u}^\kappa)) . \quad (2.14)$$

For simplification purposes, the displacement field \mathbf{u}^κ stands for the parametrization $\boldsymbol{\theta}_{T^\kappa}$ of the transformation T^κ . Thevenaz et al [34] demonstrated that the derivative of Equation (2.14) can be written in a synthetic way as

$$\frac{\partial \mathcal{S}(\mathbf{u}^\kappa + \epsilon \mathbf{h}^\kappa)}{\partial \epsilon} = \sum_{s', s} \frac{\partial p^\kappa(s', s|\mathbf{u}^\kappa + \epsilon \mathbf{h}^\kappa)}{\partial \epsilon} \cdot \log \left(\frac{p^\kappa(s', s|\mathbf{u}^\kappa + \epsilon \mathbf{h}^\kappa)}{p^\kappa(s'|\mathbf{u}^\kappa + \epsilon \mathbf{h}^\kappa)} \right) . \quad (2.15)$$

Computing the derivative regarding ϵ from Equation (2.13) yields

$$\begin{aligned} \frac{\partial p^\kappa(s', s|\mathbf{u}^\kappa + \epsilon \mathbf{h}^\kappa)}{\partial \epsilon} &= \frac{1}{\sum_{si} W_{si}} \sum_i W_{si} \cdot \beta'(\hat{D}_i^\kappa(\mathbf{x}_i + \mathbf{u}^\kappa(\mathbf{x}_i) + \epsilon \mathbf{h}^\kappa(\mathbf{x}_i)) - s') \\ &\quad \nabla \hat{D}_i^\kappa(\mathbf{x}_i + \mathbf{u}^\kappa(\mathbf{x}_i) + \epsilon \mathbf{h}^\kappa(\mathbf{x}_i)) \cdot \mathbf{h}^\kappa(\mathbf{x}_i) . \end{aligned} \quad (2.16)$$

By injecting the derivative in Equation (2.16) in Equation (2.15) evaluated at $\epsilon = 0$, one finally obtains

$$\begin{aligned} \left. \frac{\partial \mathcal{S}(\mathbf{u} + \epsilon \mathbf{h})}{\partial \epsilon} \right|_{\epsilon=0} &= \frac{1}{\sum_{si} W_{si}} \sum_i \mathbf{h}^\kappa(\mathbf{x}_i) W_{si} \nabla \hat{D}_i^\kappa(\mathbf{x}_i + \mathbf{u}^\kappa(\mathbf{x}_i)) \\ &\quad \sum_{s', s} \log \left(\frac{p^\kappa(s', s|\mathbf{u}^\kappa)}{p^\kappa(s'|\mathbf{u}^\kappa)} \right) \beta'(\hat{D}_i^\kappa(\mathbf{x}_i + \mathbf{u}^\kappa(\mathbf{x}_i)) - s') . \end{aligned} \quad (2.17)$$

It can be noted that the $\sum_{si} W_{si}$ is simply equal to the number of voxels in the image since the sum of atlas probabilities at each voxel is equal to 1. The last equation can therefore be interpreted

as a scalar product between the h^κ function and the first variation ∂_u of $\mathcal{S}(u)$ that can easily be obtained by identification

$$\partial_u \mathcal{S}(u) = W_{si} \nabla \hat{D}^\kappa(x_i + u^\kappa(x_i)) \sum_{s', s} \log \left(\frac{p^\kappa(s', s | u^\kappa)}{p^\kappa(s' | u^\kappa)} \right) \beta'(\hat{D}^\kappa(x_i + u^\kappa(x_i)) - s') \quad . \quad (2.18)$$

As described in Hermosillo et al [32] and Rougon et al [33], the variational optimization of \mathcal{S} is solved by looking for the displacement field that cancels the derivative in Equation (2.11). Such local minimum can be reached by iteratively composing the first variation given by Equation (2.18) with the current estimate of the displacement field. The W_{si} factor in Equation (2.18) weights the contribution of each label by the priority to observe the label at this point in the existing reference. Moreover, the presence of $\nabla \hat{D}^\kappa$ reminds a classical optical flow incremental field estimation. Instead of being multiplied by the difference of intensity functions in fixed and moving images, the gradient in the subject (moving image) is multiplied by a weighted function of $W_{si} \cdot \beta'(\cdot)$. The weights are proportional to the joint and marginal probability densities measured over the entire atlas domain. This proves that mutual information (in opposition to mean squared differences) acts as a global metric, even for estimating local displacements, taking into account global statistical properties of both modalities to be aligned.

2.3.4. Iterative scheme

To accelerate the convergence of the method for practical experiments, the transformation parameters are first kept constant. For this configuration, iterating between Equations (2.2) and (2.6) enables us to converge to a stable W_{si} reference (that is, the estimate of the probabilistic atlas for the existing set of transformations).

Further, each subject is aligned on the current W_{si} probabilistic atlas by maximizing the criterion in Equation (2.9) for an affine transformation model as described in Section 2.3.2.

Once the atlas has reached convergence, this entire iterative scheme is repeated for a non-parametric representation of the non-rigid atlas to subject's transformation (resulting in a dense deformation field) as described in Section 2.3.3.

3. Results

The results are presented in three sections. In the first section, our group-wise alignment algorithm has been applied to a database of statistically significant size (80 subjects) for affine and non-rigid (dense field) transformation models. The results are assessed visually by plotting an overlay of 9 subjects with the contours of the final atlas estimate and by plotting the sum of the 80 segmentations following rigid and non-rigid alignment. For non-rigid transformations, regularity of the displacement fields is ensured by positiveness of the transformation Jacobians. The second section examines the stability of the atlas construction process when new subjects are added to the

Affine		
metric	number of spatial samples	10000
metric	PDF B-Spline order	3
SPSA	number of iterations	500
SPSA	first step magnitude	0.5
SPSA	finite difference initial step	0.5
SPSA	number of stochastic gradient averages	4
Non-rigid		
metric	number of spatial samples	10000
metric	PDF B-Spline order	3
pre-processing	standard deviation for initial Gaussian image smoothing	3 mm
optimization	number of iterations	50
optimization	learning rate	2.5e4
regularization	standard deviation for Gaussian smoothing of disp. field	3.5 mm

Table 1: Set of parameters used for affine and non-rigid group-wise registration. See Section 3.

database. The next section compares quantitatively the representativeness of our group-wise reference with the classical choice of individuals in the population as reference. The representativeness of a reference with a given population is measured using William’s index [35]. Finally, the last section measures the ability of our non-rigid group-wise algorithm to recover synthetic deformations. The accuracy of the alignment process is measured using a global overlap coefficient similar to the one Zollei [8, 9] has described. The robustness of the algorithm when noisy subjects are incorporated to the population is also examined in this section.

The set of parameters that we use for group-wise affine and non-rigid alignment is given in Table 1. The first type of parameters, common to affine and non-rigid alignment, configures the similarity metric between the atlas and each subject. The number of spatial samples is the number of random samples taken in the atlas domain to estimate the joint probability in Equations 2.10 and 2.13. The order of the B-Spline basis functions for estimating the joint probability density functions is set to 3, which is equal to the number of bins affected by each sample in the joint histogram. The total number of bins in the joint histogram is equal to the number of labels in the segmented data. The next set of parameters is related to the optimization process. In the affine registration case, all optimization parameters are related to the SPSA optimizer. The number of iterations is a fixed number, corresponding to the number of steps in the gradient directions. At each iteration, the gradient is estimated by averaging 4 stochastic estimates. This has the advantage to stabilize the stochastic optimization process compared to using one stochastic perturbation only [30]. Each perturbation is performed simultaneously on all parameters as described in Spall et al [30]. The magnitude of the perturbation, as well as the learning rate, is decreasing in an exponential way over the iterations. The coefficients α and γ [30] describing this decay are $\alpha = 0.602$ (learning rate decrease) and $\gamma = 0.101$ (perturbation amplitude decrease), which are common choices when performing SPSA optimization [30]. For the non-rigid registration, the most important parameters

are related to the regularization of the displacement field. At each iteration, the displacement field is regularized using a Gaussian smoothing with a standard deviation given in the table above. The input images are regularized as well using a Gaussian filter.

3.1. Alignment of 80 brain segmentations

The group-wise registration algorithm has been run on 80 segmentations of adult brains.

The population is made of adult schizophrenic subjects who underwent Magnetic Resonance imaging. All MR images have the same dimension ($256 \times 256 \times 124$ voxels) and image resolution ($0.9375 \times 0.9375 \times 1.5$ millimeters). The whole dataset has been manually segmented using the Slicer software (<http://www.slicer.org>) in four labels : background, CSF, white and gray matter. The skull was not labelled in this operation and is assimilated to the background class.

The set of group-wise transformations has been initialized by translations bringing the center of gravity of each subject (computed on the mask of the intra-cranial cavity by averaging all coordinates inside the mask) to the geometrical center of the atlas domain.

3.1.1. Affine Alignment

Two iterations, each of them including atlas estimation and sequential subject to atlas alignment, were used to find the optimal parameters of an affine transformation model parametrized by 9 parameters. The scaling and rotation matrices are applied in the following order: rotations around Z, Y and X axis, followed by a diagonal matrix for scaling X, Y and Z axis. Translations are applied after rotation and scaling. The algorithm for this step is described in Section 2.3.2. In our transformation model, the scaling parameters are mapped to the $(-\infty, +\infty)$ domain by optimizing their logarithm (this avoids to have different search domains for each parameter). The alignment of a subject on the current atlas estimate takes less than 5 minutes. The SPSA optimization routine and the mutual information metric for probabilistic atlas to image registration have been integrated in the Insight Toolkit[36].

3.1.2. Non-Rigid Alignment

For non-rigid registration, the flow in Equation 2.18 has been integrated in the ITK[36] dense multi-threaded non-rigid registration framework. The non-rigid alignment of one subject on the current reference estimates takes about 50 minutes on a 64 bits 2GHz CPU. This time needs to be multiplied by the number of subjects to obtain the total non-rigid alignment time. With our database, we performed 2 times the whole cycle of atlas estimation and non-rigid alignment on this reference. This makes the total computation time for the non-rigid part close to 135 hours. Hopefully, this time could be considerably reduced by running the algorithm on multi-threaded computers. The non-rigid alignment of the database on the atlas could be considerably reduced by distributed computing on a grid architecture since the alignment of each subject can be performed independently.

The displacement field is smoothed at each iteration by a Gaussian kernel with a standard deviation given in Table 1. Figure 1 plots a simple unweighted sum of the labels across all subjects

Subject	Min. Jaco.	Max. Jaco.	Mean Jaco.
case 1/80	0.765	1.544	1.012
case 3/80	0.731	1.681	1.014
case 10/80	0.679	1.877	1.008
case 15/80	0.750	1.354	1.015
case 13/80	0.829	1.426	1.019
case 23/80	0.733	1.622	1.012
case 71/80	0.716	1.725	1.010
case 25/80	0.818	1.710	1.019
case 45/80	0.825	1.693	1.012

Table 2: Minimum, maximum and mean Jacobian values obtained by group-wise non-rigid transformation of the height subjects plotted in Figures 4, 5, 2 and 3. The absence of negative values guarantees that the non-rigid transformations of these subjects are invertible.

following rigid (left) and non-rigid (right) alignment. The contours of the ventricles in the sum after non-rigid alignment appear more precisely after non-rigid alignment.

To qualitatively assess the efficiency of the rigid and non-rigid group-wise registration algorithms, Figures 2 and 3 plot the same 9 subjects following affine and non-rigid alignment for a sagittal slice. Figures 4 and 5 depict the same subjects as Figures 2 and 3 but in a coronal view. These figures show that the non-rigid transformations recovered by our algorithm are effectively attracting the external contours of the brains and ventricles towards the atlas contours.

To determine if the non-rigid deformation fields produced by our algorithm preserve the topology inside the brain, one needs to compute the range of values taken by the transformation Jacobian. Negative Jacobian values indicate that the displacement field either produces folding or tends to remove or create entire regions in the image. To ensure that our algorithm produces invertible transformations, Table 2 shows the minimum, maximum and mean Jacobian values inside the image volume for the same eight subjects as in Figures 2, 3, 4 and 5. It can be verified that no negative value is reported, the minimal Jacobian value for these subjects being greater than 0.7.

Figure 1 also depicts that some small clusters of white matter that appear more intense following non-rigid alignment. This suggests that the number of coherent white matter pixels across all subjects increases with the non-rigid alignment of these subjects on the probabilistic atlas.

3.2. Stability of the atlas construction procedure for incrementing population sizes

When evaluating a group-wise registration algorithm, it is important to examine how an atlas evolves when new subjects are added to the population. An intuitive and expected behavior when considering successive extensions of the population is to observe a decreasing variation of the atlas both in value and rate. The variation of the reference for incremental population sizes should therefore show a decreasing and asymptotically converging curve.

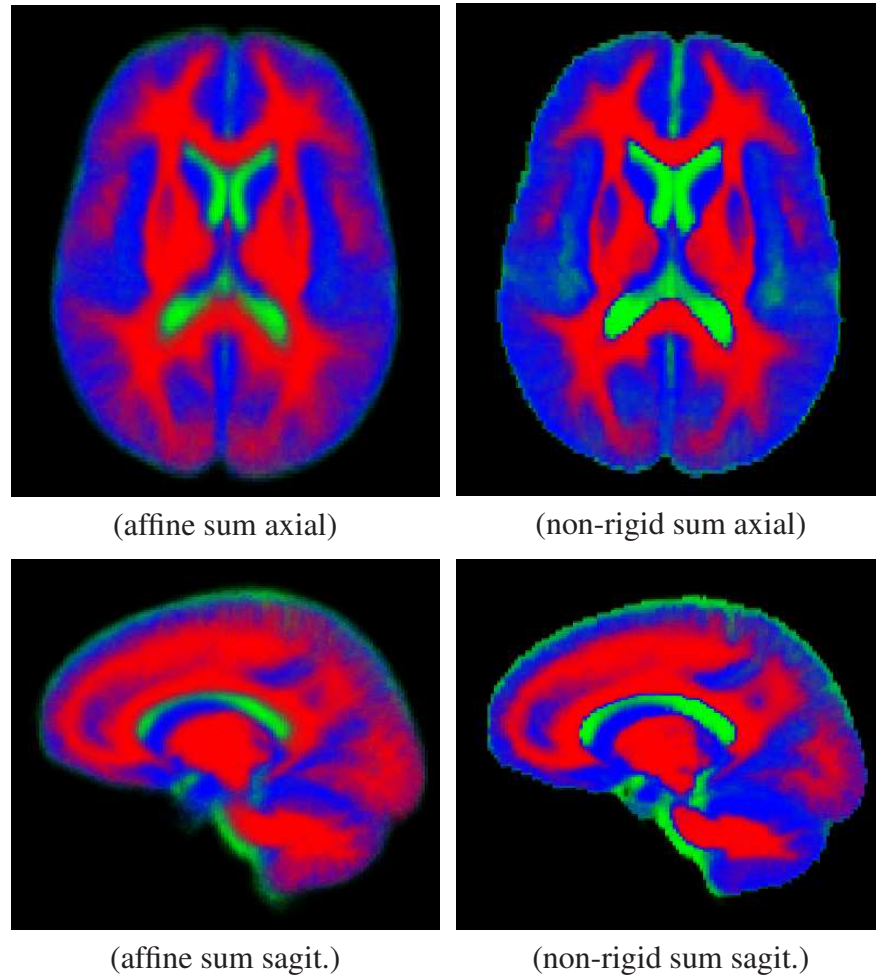


Figure 1: Sum of the 80 segmentations (four labels) after affine alignment and after 2 iterations of non-rigid warping. The value plotted at each pixel is the sum of aligned segmentations ($\sum_j D_j(T_j(i))$) with each label plotted in a different channel. The contours around the ventricles and the external borders are less fuzzy after non-rigid alignment which shows that the uncertainty in these zones is reduced compared to an affine alignment. The white matter appears more compact following non-rigid alignment i.e. in the cerebellum (dentate nucleus).

In this experiment, the algorithm was run for affine transformations (Section 2.3.2.) for populations that increased incrementally from 10 to 75 subjects. This affine alignment is performed independently for each incremental size of population. Groups of 5 subjects are added as they appear in alphabetical order among the subject file names. The resulting W_{si} atlas for each population size was compared to the final atlas by computing the root mean squared error (RMS) over all voxels for each label between the atlas estimate and the final W_{si} map obtained when aligning all 80 subjects in the database.

$$RMS_s = \left(\frac{1}{N_{voxels}} \cdot \sum_i (W_{si}^{current} - W_{si}^{final})^2 \right)^{1/2} \quad (3.1)$$

Figure 6 shows the evolution of the per label RMS for incrementing sizes of the population. The averaged RMS over all labels, defined as

$$RMS_{AV} = \left(\frac{1}{N_{voxels} \cdot N_{labels}} \cdot \sum_{i,s} (W_{si}^{current} - W_{si}^{final})^2 \right)^{1/2} \quad (3.2)$$

is plotted in Figure 7 for incrementing population sizes from 10 to 75 subjects.

The global decrease of the RMS curve shows how, as the size of the population increases, the reference for aligning the population becomes more stable when new subjects are added. This is mainly due to the computation of the W_{si} map (see Equation (2.2)) which is not a linear mean of subject labels, and takes into account the subject similarity parameters for the estimated alignment.

3.3. Comparison of the EM reference representativeness with a pairwise registration strategy

In comparing our algorithm to existing pairwise strategies, the reference W_{si} generated by the group-wise registration algorithm has to be matched with the choice of an arbitrary subject in the database as a reference. To this end, we arbitrarily selected a subset of 10 subjects in the population. Each of these 10 subjects was in turn considered as a reference and the 9 other subjects were aligned on the reference using a pairwise metric and an affine transformation model. The pairwise similarity metric used in this experiment is the *MatchCardinalityImageToImageMetric* provided by the ITK Insight Toolkit. This metric computes the cardinality of the set of pixels that match exactly between the moving and fixed images. The magnitude of an individual label mismatch is not relevant, however; only the occurrence of a label mismatch is important. Previously this metric has been used for alignment of segmentations from medical images [37] and has been demonstrated to be extremely robust and very fast when implemented to run in parallel on shared memory and distributed memory machines.

Each of the 10 choices of reference generates 10 aligned populations. For each of these group-wise alignment configurations, we measured the agreement between each reference and the aligned population using William's index [35]. This index measures the agreement of a reference to each

subject of a population compared with the mean agreement inside the population. The definition of William's index for an agreement operator $a(\cdot)$ is

$$\text{Index}_{ref} = \frac{(N_{\text{pop}} - 1) \sum_j^{N_{\text{pop}}} a(S_{ref}, S_j)}{2 \sum_j^{N_{\text{pop}}} \sum_{j'}^{j-1} a(S_j, S_{j'})} \quad (3.3)$$

where N_{pop} stands for the number of subjects in the population (without counting the reference subject) and S_j or $S_{j'}$ designate subjects of the population.

Different agreement measures can be used for quantifying the similarity between two masks [35]. In our case, we worked with the Jaccard measure, which compares the cardinality of two masks intersection over the cardinality of their union

$$JC = \frac{|X \cap Y|}{|X \cup Y|} \quad (3.4)$$

Table 3 shows for each choice of reference, the William's index value of this reference regarding the aligned population. The index is computed for each label separately since this measure cannot handle multiple labels. Table 3 shows that the larger variation among the measured indices is observed for the CSF label (ventricles). It can be seen that following William's index, our group-wise reference is more representative than any choice of pairwise reference in this population. The last 4 columns show William's index obtained by a reference generated using a majority voting rule. Even if this reference scheme outperforms our group-wise reference in the sense of William's index, it generates noisy images since decisions at neighboring are independent as shown in Figure 8.

Because of the pretty small size of the population considered in this experiment, the non-rigid group-wise registration algorithm described in Sect. 2.3.2. was run for comparison on the 10 populations obtained by pairwise registration. This way, one can effectively compare the representativeness of the pairwise and the group-wise references for the same set of nine subjects. For each population of nine subjects, William's index is computed between the W_{si} atlas and the 9 aligned brains. In applying Equation 3.3, the reference label at each voxel is taken as the label having the maximal probability in the atlas. The result for each label is reported in Table 3. It can be seen that the EM reference produced by our algorithm is more concordant with the population than the subject taken as reference for affine alignment. For comparison purposes, we generated, for the 10 populations obtained by pairwise affine alignment, an average obtained by selecting at each voxel the more represented label in the population. William's index between this average and all nine subjects was then computed for each population. Results given in Table 3 show that this "majority vote" outperforms our group-wise reference. To understand the reason behind this, all 10 references obtained by majority votes in the 10 populations have been rigidly aligned for visualization purpose and plotted in Figure 8. Several inconsistencies can be observed when applying the majority rule in areas of higher anatomical variability. Since the decisions at all voxels are independent, contradictory label can be elected at neighboring voxels. In comparison, the group-wise reference produced by our algorithm is obtained by weighting all subjects in a coherent way across all voxels. The resulting labelling as shown in the bottom right of Figure 8 (plotting at each voxel the label

with the highest probability) is less noisy than all references obtained by majority vote. This can be explained by the fact that our group-wise reference weights the different subjects of the population consistently at all voxels, since the coefficients of this rating are based on the specificity/sensitivity performances of each subject. The higher scores obtained using William’s index by the majority reference scheme therefore show that this index is biased towards the decision of the majority at each voxel, regardless of the consistency of these decisions across neighbouring voxels.

ref. number	Pairwise				Group-wise				Majority votes			
	back.	CSF	GM	WM	back.	CSF	GM	WM	back.	CSF	GM	WM
ref. 1	1.001	1.093	1.014	1.013	1.005	1.601	1.203	1.231	1.006	1.720	1.401	1.292
ref. 2	1.002	1.048	1.028	1.037	1.004	1.546	1.183	1.207	1.006	1.711	1.406	1.296
ref. 3	1.001	1.060	1.052	0.965	1.005	1.588	1.195	1.220	1.006	1.711	1.407	1.291
ref. 4	1.003	0.701	1.051	1.058	1.005	1.611	1.200	1.232	1.005	1.680	1.402	1.295
ref. 5	1.001	1.239	1.008	1.040	1.005	1.604	1.205	1.228	1.005	1.711	1.402	1.294
ref. 6	1.000	1.174	1.032	1.004	1.005	1.596	1.184	1.213	1.006	1.723	1.408	1.295
ref. 7	1.001	1.062	1.008	0.997	1.005	1.629	1.215	1.227	1.006	1.717	1.400	1.290
ref. 8	0.996	0.988	0.885	0.919	1.005	1.621	1.193	1.223	1.005	1.709	1.392	1.289
ref. 9	1.001	1.106	1.015	1.038	1.005	1.629	1.213	1.237	1.006	1.735	1.402	1.295
ref. 10	1.000	0.919	1.041	1.048	1.005	1.583	1.201	1.230	1.007	1.713	1.405	1.296

Table 3: William’s indexes comparison between a pairwise reference choice and group-wise reference for a population of 10 subjects for background (back), Cerebrospinal Fluid (CSF), Gray Matter (GM) and White Matter(WM) labels. The first 4 columns of this table show the William’s index value between each subject taken as a reference and the 9 other subjects aligned on the reference using an transformation model. The next 4 columns show the values of William’s index measured between the group-wise reference produced by our algorithm and the same population aligned on this reference using an affine transformation model as described in Section 2.3.2. The last 4 columns show the same index when taking a majority vote rule as reference.

3.4. Synthetic non-rigid transformations

In this section, we first examine how our algorithm can align a collection of 10 subjects synthetically generated by a B-Spline deformation model from one single subject. The coefficients of the B-Spline transformations are generated using a Gaussian distribution. The random number generator used in this experiment is the NormalVariateGenerator provided in the ITK toolkit [36]. The mean of the Gaussian distribution was set to 0 and the standard deviation to 4 millimeters (the original dataset has a spacing of $0.9375 \times 0.9375 \times 1.51 \text{ mm}^3$).

The non-rigid group-wise registration algorithm that has been described in Section 2.3.3. was run on the synthetic set of 10 subjects. For showing numerically how the population gets into alignment, we used a global overlap measure introduced by Zollei [9]

$$\text{Overlap Coefficient} = \frac{|A_1 \cap \dots \cap A_{10}|}{\min(|A_1|, \dots, |A_{10}|)} \quad (3.5)$$

Figure 9 plots this overlap coefficient over consecutive iterations of the group-wise non-rigid alignment algorithm described in Section 3.1.2. The blue bars plot, for each label, the initial overlap coefficient after applying the 10 random B-Spline transformations. The green and red bars plot the value of the overlap coefficient after consecutive passes of the algorithm (probabilistic reference estimation using Equation (2.2) and alignment of all subjects on the reference using the metric flow defined in Equation (2.18)). The poor values of the overlap coefficient for the CSF label can be explained by the high sensitivity of Eq. 3.5 for small anatomical structures: if in one voxel, one subject only is in disagreement with the others, the measure will return 0 for this point.

Next, for examining the robustness of the reference computation scheme regarding the presence of outliers in the database, noisy subjects are added to the initial 10 subjects. These noisy subjects are produced in two steps. First, the original subject was deformed using a random B-Spline transformation like in the previous experiment. A “pepper and salt” noise was then added by replacing with a fixed probability (0.2 in our case) all voxels to an arbitrary label. Such perturbation is performed to estimate the robustness of our algorithm to outliers in terms of segmentation quality.

Since the reference computation scheme (Equation (2.2)) takes into account the consistency of each subject in the reference scheme, the reference is only slightly disturbed, even when adding a significant number of noisy subjects (5 for an initial population of 10 subjects in this case). To validate this statement, the initial population of 10 synthetic subjects has been augmented by 3 and then 5 noisy subjects. This new population was aligned using the non-rigid group-wise alignment algorithm of Section 2.3.3. The influence of the addition of the outliers was measured by the variation of the global overlap index after two passes of our non-rigid group-wise registration algorithm. Figure 10 shows the variation of this index for 0, 3 and 5 outliers. It can be observed that the performances of our group-wise registration algorithm are not significantly altered after the introduction of outliers in the initial population.

4. Discussion and Conclusion

In this paper, we have proposed to address the group-wise registration problem within the Expectation-Maximization framework. We have shown that in using this framework, the group-wise registration of a population of segmented objects can be divided into two steps: estimation of a reference standard and alignment of each subject in the population to this reference. In our case, the reference is represented as a probabilistic atlas giving at each voxel, the probability of observing a tissue class at this point after alignment of the population. This reference standard then defines a common coordinate system for the population.

This reference scheme provides answers to the two questions that were framed in the introduction: First, can we define a common space of coordinates that is unbiased by choosing an arbitrary subject from the population reference ? Secondly, how can we deal robustly with the presence of outliers in the population that could disturb a reference computation scheme based on the arithmetic mean (e.g. [18]).

With respect to the former question, we can only achieve unbiased construction by utilizing a combination of *all* subjects for the available alignment instead of choosing an arbitrary subject as a reference. With respect to the latter question, the reference scheme weights each subject based on the similarity parameters of this subject with the estimated reference. We have illustrated the advantages of our reference scheme against pairwise reference choices as classically performed by Guimond et al [20] and Rueckert et al [21]. We also have shown that our reference was consistent across neighbouring voxels since it weights all subjects consistently depending on their specificity/sensitivity performances. This shows a clear advantage in comparison with reference computation based on averaging the population [18] or a voting approach. The main specificity of our approach, compared to existing frameworks, is that both the atlas and the group-wise alignment proceed from the optimization of the same cost function, the reference being modeled as hidden variable of the Expectation-Maximization problem.

Even if we show that the reference stabilizes for large populations when new subjects are added to the database (Section 3.2.), the algorithm only computes a single reference and does not yet permit us to differentiate among clusters in the population. Future work will estimate the number of clusters that will be integrated into the current estimation scheme and will further assess a given subject's probability of belonging to each cluster.

Another disadvantage of the method is that it deals with labeled images. Exploring how to adapt the method to gray level images may be a line of research in the future.

Acknowledgements

This work was supported by the Belgian Walloon Region by a First Europe grant (Convention EPI A320501R049F /415732), by NSF ITR 0426558, a research grant from CIMIT, grant RG 3478A2/2 from the NMSS, and by NIH grants R21 MH067054, R01 RR021885 and U54 EB005149.

References

- [1] J. Talairach, P. Tournoux, Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging. Thieme Medical Publishers, New York, 1988.
- [2] R. Kikinis, M. E. Shenton, D. V. Iosifescu, R. W. McCarley, P. Saiviroonporn, H. H. Hokama, A. Robatino, D. Metcalf, C. G. Wible, C. M. Portas, R. Donnino, F. A. Jolesz, *A digital brain atlas for surgical planning, model driven segmentation and teaching*, IEEE Transactions on Visualization and Computer Graphics, 2 (1996), No. 3, 232–241.
- [3] B. Dawant, S. Hartmann, J.-P. Thirion, F. Maes, D. Vandermeulen, P. Demaerel, *Automatic 3-D segmentation of internal structures of the head in MR images using a combination of similarity and free-form transformations. Part I. Methodology and validation on normal subjects*, IEEE Transactions on Medical Imaging, 18 (1999), No. 10, 909–916.

- [4] K. M. Pohl, J. Fisher, E. Grimson, R. Kikinis, W. M. Wells, *A bayesian model for joint segmentation and registration*, NeuroImage, 31 (2006), No. 1, 228–239.
- [5] T. Rohlfing, D. Russakoff, C. Maurer, *Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation*, IEEE Transactions on Medical Imaging, 23 (2004), No. 8, 983–994.
- [6] C. Studholme, V. Cardenas, *A template free approach to volumetric spatial normalization of brain anatomy*, Pattern Recognition Letters, 25 (2004), No. 10, 1191–1202.
- [7] K. Bhatia, J. Hajnal, B. Puri, A. Edwards, D. Rueckert, *Consistent groupwise non-rigid registration for atlas construction*, in IEEE International Symposium on Biomedical Imaging, 2004, 908–911.
- [8] L. Zollei, E. Learned-Miller, E. Crimson, W. Wells, *Efficient Population Registration of 3D Data*, in Computer Vision for Biomedical Image Applications: First International Workshop, 2005.
- [9] L. Zollei, *A unified information theoretic framework for pair- and group-wise registration of medical images*, Ph.D. dissertation, Massachusetts Institute of Technology Computer Science and Artificial Intelligence Laboratory, 2006.
- [10] S. Marsland, C. Twining, C. Taylor, *A minimum description length objective function for groupwise non-rigid image registration*, Image and Vision Computing, 26 (2008), No. 3, 333–346.
- [11] D. C. Van Essen, *A population-average, landmark- and surface-based (pals) atlas of human cerebral cortex*. Neuroimage, 28 (2005), No. 3, 635–662.
- [12] M. Miller, S. Joshi, G. Christensen, Brain Warping. Academic Press, San Diego, 1999, chapter 7, *Large deformation fluid diffeomorphisms for landmark and image matching*, 115–131.
- [13] S. Dinggang, C. Davatzikos, *Hammer, hierarchical attribute matching mechanism for elastic registration*, IEEE Transactions on Medical Imaging, 21 (2002), No. 11, 1421–1439.
- [14] D. L. Collins, P. Neelin, T. M. Peters, A. C. Evans, *Automatic 3d intersubject registration of MR volumetric data in standardized Talairach space*. Journal of Computer Assisted Tomography, 18 (1994), No. 2, 192–205.
- [15] J. Mazziotta, *A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM)*, Philosophical Transactions: Biological Sciences, 356 (2001), No. 1412, 1293–1322.
- [16] J. Csernansky, S. Joshi, L. Wang, J. Haller, M. Gado, J. Miller, U. Grenander, M. Miller, *Hippocampal morphometry in schizophrenia by high dimensional brain mapping*, Proceedings of the National Academy of Sciences, 95 (1998), No. 19, 11406.

- [17] P. Thompson, J. Moussai, S. Zohoori, A. Goldkorn, K. A. A., M. Mega, G. Small, J. Cummings, A.W. Toga, *Cortical variability and asymmetry in normal aging and alzheimer's disease*, Cerebral Cortex, 8 (1998), No. 6, 492–509.
- [18] S. Joshi, B. Davis, M. Jomier, G. Gerig, *Unbiased diffeomorphic atlas construction for computational anatomy*, Neuroimage, 23 (2004), 151–160.
- [19] J. Diedrichsen, *A spatially unbiased atlas template of the human cerebellum*, Neuroimage, 33 (2006), No. 1, 127–138.
- [20] A. Guimond, J. Meunier, J.-P. Thirion, *Average brain models: A convergence study*, Computer Vision and Image Understanding, 77 (1999), No. 2, 192–210.
- [21] D. Rueckert, A. F. Frangi, J. A. Schnabel, *Automatic construction of 3D statistical deformation models of the brain using nonrigid registration*, IEEE Transactions on Medical Imaging, 22 (2003), No. 8, 1014–25.
- [22] V. Arsigny, O. Commowick, X. Pennec, N. Ayache, *A Log-Euclidean framework for statistics on diffeomorphisms*, in Medical Image Computing and Computer-Assisted Intervention, LNCS 4190 (2006), 924–931.
- [23] C. Twining, S. Marsland, *Constructing an atlas for the diffeomorphism group of a compact manifold with boundary, with application to the analysis of image registrations*, Journal of Computational and Applied Mathematics, In Press, Corrected Proof, 2008.
- [24] S. Marsland, C. Twining, C. Taylor, *Groupwise non-rigid registration using polyharmonic clamped-plate splines*, in Medical Image Computing and Computer-Aided Intervention, LNCS 2879 (2003), 771–779.
- [25] P. Lorenzen, M. Prastawa, B. Davis, G. Gerig, E. Bullitt, S. Joshi, *Multi-modal image set registration and atlas formation*, Medical Image Analysis, 10 (2006), No. 3, 440–451.
- [26] S. Warfield, K. Zou, W. Wells, *Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation*, IEEE Transactions on Medical Imaging, 23 (2004), No. 7, 903–921.
- [27] A. Roche, X. Pennec, G. Malandain, N. Ayache, *Rigid registration of 3D ultrasound with MR images: a new approach combining intensity and gradient information*, IEEE Transactions on Medical Imaging, 20 (2001), No. 10, 1038–1049.
- [28] P. Viola, W. M. Wells, *Multi-modal volume registration by maximization of mutual information*, Medical Image Analysis, 1 (1996), No. 1, 35–51.
- [29] D. Mattes, D. Haynor, H. Vesselle, T. Lewellen, W. Eubank, *Pet-ct image registration in the chest using free-form deformations*, IEEE Transaction on Medical Imaging, 22 (2003), No. 1, 120–128.

- [30] J. Spall, *Adaptive stochastic approximation by the simultaneous perturbation method*, IEEE Transactions on Automatic Control, 45 (2000), 1839–1853.
- [31] A. Cole-Rhodes, K. Johnson, J. LeMoigne, I. Zavorin, *Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient*, IEEE Transactions on Image Processing, 12 (2003), 1495–1511.
- [32] G. Hermosillo, C. Chefd’Hotel, O. Faugeras, *Variational methods for multimodal image matching*, International Journal of Computer Vision, 50 (2002), No. 3, 329–343.
- [33] N. Rougon, C. Petitjean, F. Preteux, P. Cluzel, P. Grenier, *A non-rigid registration approach for quantifying myocardial contraction in tagged mri using generalized information measures*, Medical Image Analysis, 9 (2005), No. 4, 353–375.
- [34] P. Thevenaz, M. Unser, *Optimization of mutual information for multiresolution image registration*, IEEE Transactions on Image Processing, 9 (2000), No. 12, 2083 – 2099.
- [35] M. Martin-Fernandez, S. Bouix, L. Ungar, R. McCarley, M. Shenton, *Two methods for validating brain tissue classifiers*, in Medical Image Computing and Computer-Aided Intervention, LNCS 3749 (2005), 515–522.
- [36] L. Ibanez, W. Schroeder, L. Ng, J. Cates, *The ITK software guide*, 2nd ed., Kitware, Inc. ISBN 1-930934-15-7, 2005.
- [37] S. K. Warfield, F. A. Jolesz, R. Kikinis, *A high performance computing approach to the registration of medical imaging data*. Parallel Computing, 24 (1998), 1345–1368.

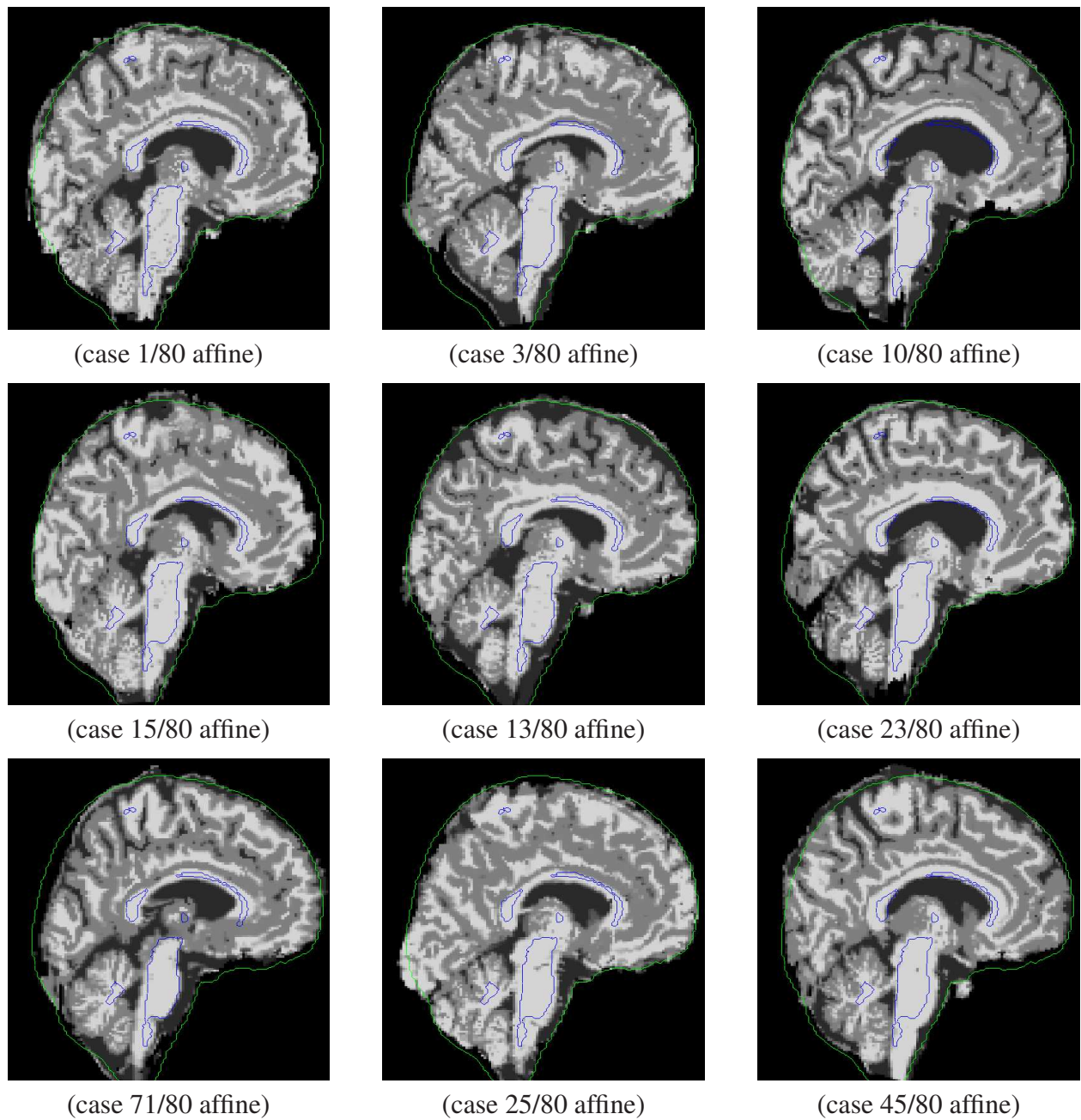


Figure 2: Affine alignment of typical subjects (sagittal slice). This figure shows the alignment of 9 subjects (chosen randomly in the 80 subjects) after affine alignment. The contours of the resulting probabilistic atlas are shown in blue (white matter border) and red (brain-background border). Switching between Figures 2 and 3 show how these subjects convert to a central tendency after non-rigid alignment.

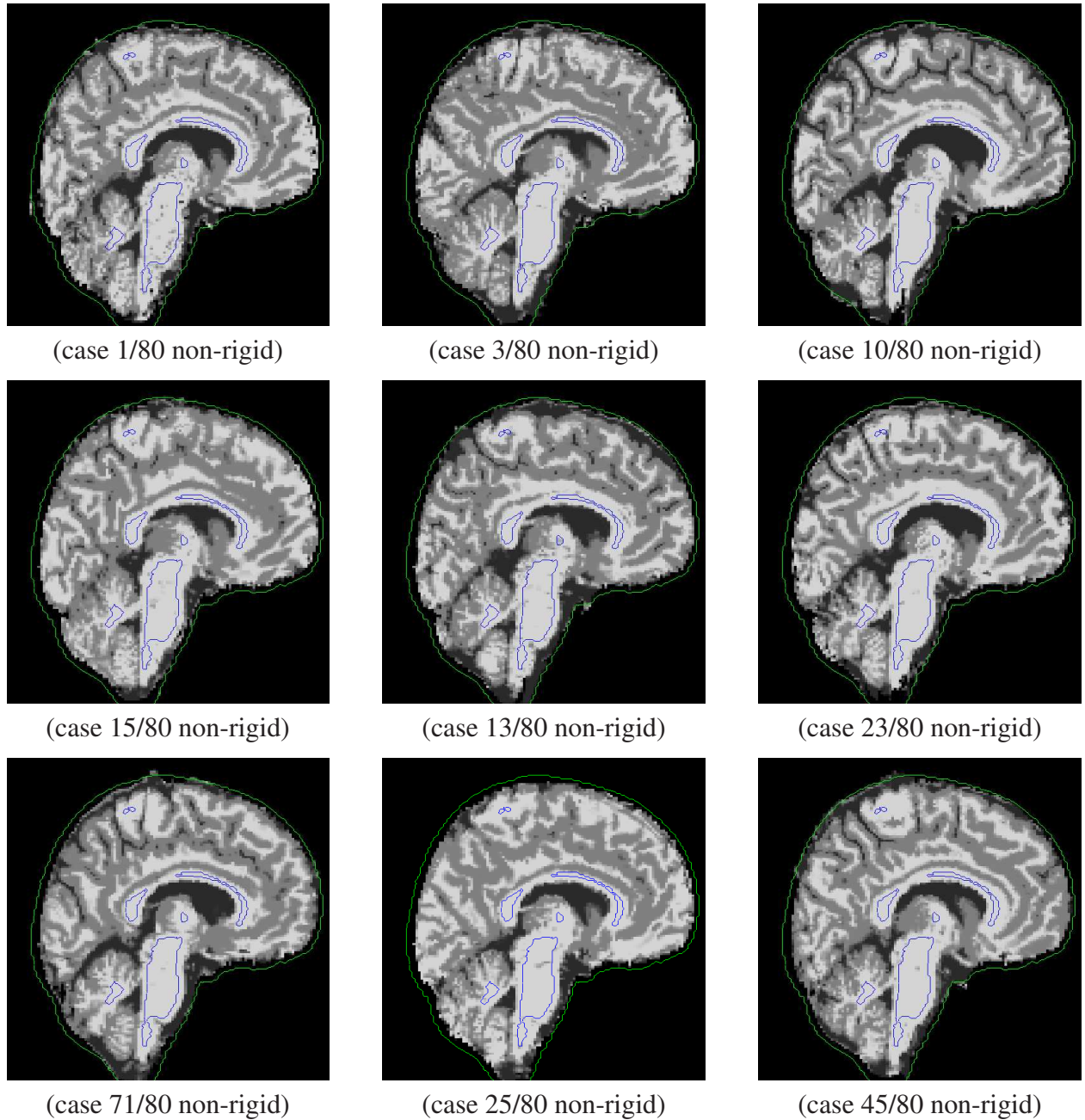


Figure 3: Non-rigid alignment of typical subjects (sagittal slice). This figure shows the alignment of the same 9 subjects as in Figure 2 after non-rigid alignment. The contours of the resulting probabilistic atlas are shown in blue (white matter border) and red (brain-background border). Switching between Figures 2 and 3 show how these subjects convert to a central tendency after non-rigid alignment. An animation for making visible changes from Figure 2 can be downloaded from http://www.cilab.upf.edu/~mdecraene/atlasanim/anim_sag.gif

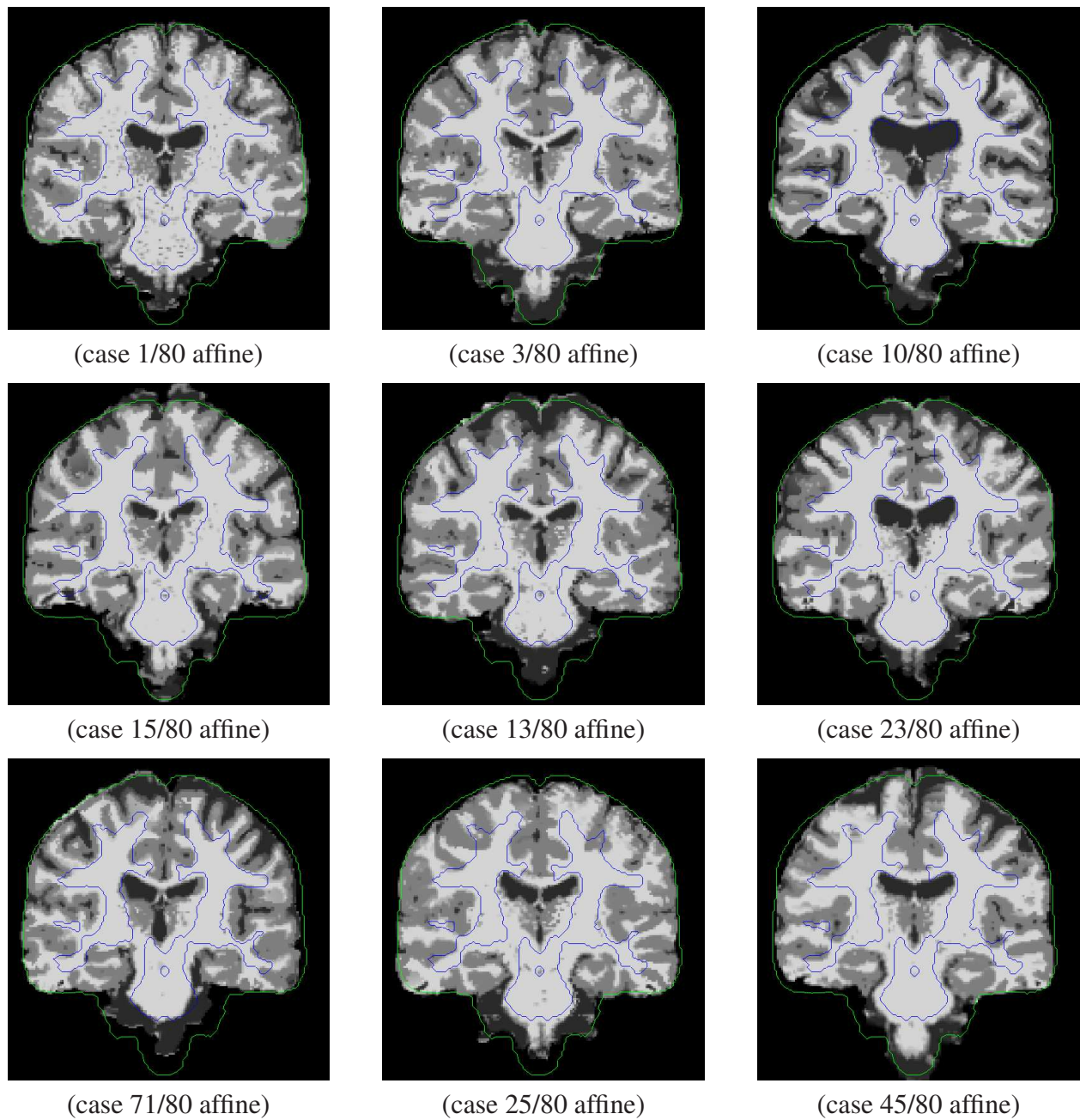


Figure 4: Affine alignment of typical subjects (coronal slice). This figure shows the affine alignment of the same subjects as in Figures 2 and 3 on a coronal slice. The contours of the resulting probabilistic atlas are shown in blue (white matter border) and red (brain-background border). Switching between Figures 4 and 5 show how these subjects convert to a central tendency after non-rigid alignment.

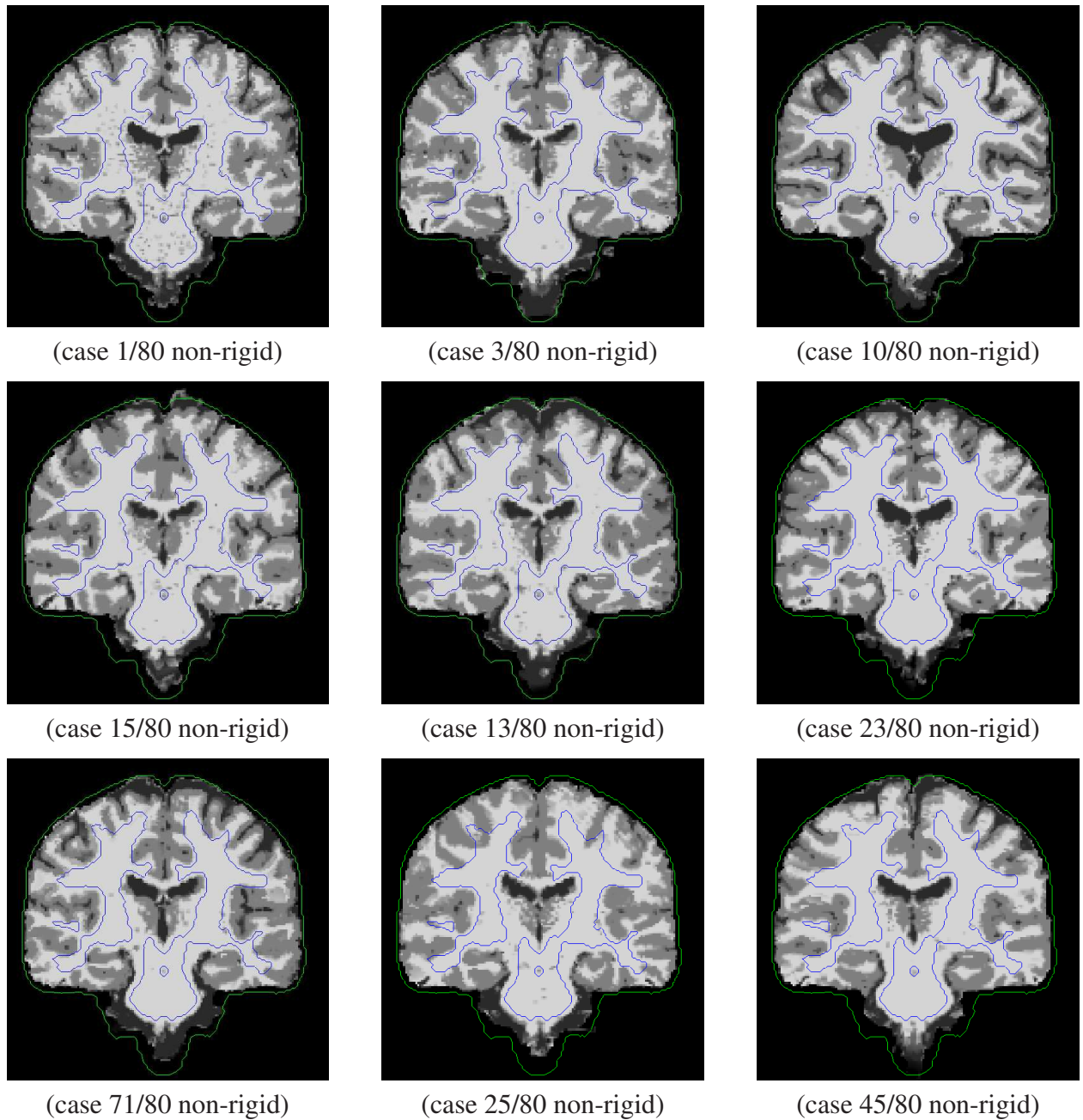


Figure 5: Non-rigid alignment of typical subjects (coronal slice). This figure shows the non-rigid alignment of the same subjects as in Figures 2 and 3 on a coronal slice. The contours of the resulting probabilistic atlas are shown in blue (white matter border) and red (brain-background border). Switching between Figures 4 and 5 show how these subjects convert to a central tendency after non-rigid alignment. An animation making visible changes from Figure 4 can be downloaded from http://www.cilab.upf.edu/~mdecraene/atlasanim/anim_cor.gif

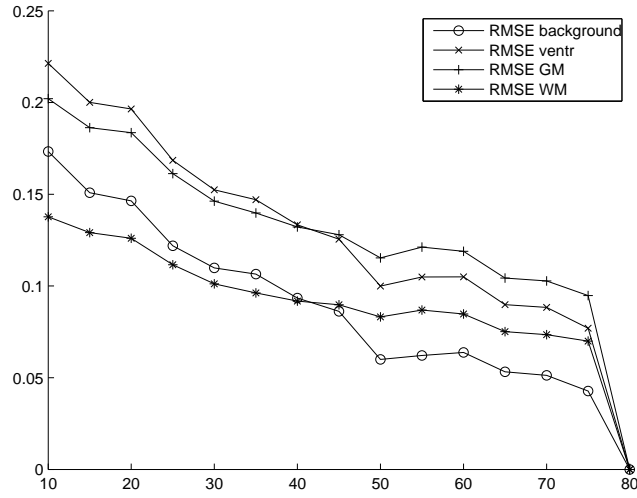


Figure 6: Plot of the RMS error between the W_{si} atlas for incremental sizes of populations and the W_{si} atlas obtained after alignment of 80 subjects. Each curve plots Eq. (3.1) for each label s . The decreasing in the slope of the curve shows that the atlas estimate gets more stable when the size of the population increases.

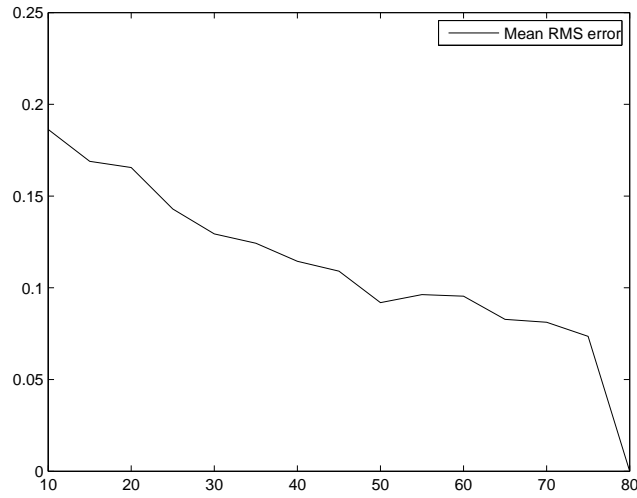


Figure 7: Mean RMS error difference (Eq. (3.2)) over all labels between the W_{si} for incremental sizes of populations and the W_{si} atlas obtained after alignment of 80 subjects. The RMS decrease per label is shown in Fig. 6

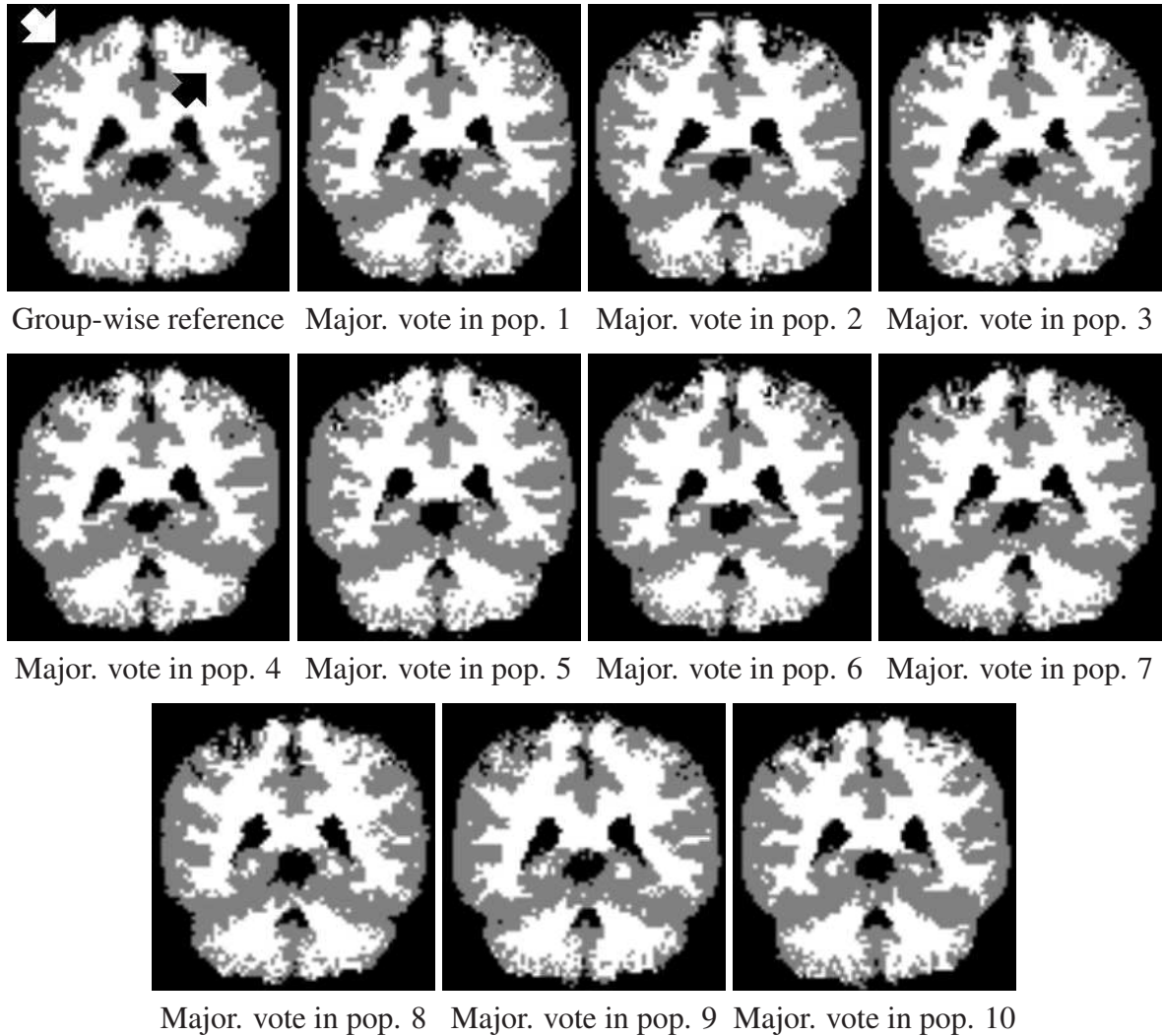


Figure 8: Comparison of references obtained by majority vote in 10 populations generated by pairwise registration and the reference generated by our group-wise registration algorithm. It can be observed that a reference obtained by majority vote at each voxel is corrupted by noise since decisions at neighboring are completely independent. In contrast, the group-wise reference produced by our algorithm shows a topologically correct labelling, consistent even in areas of high anatomical variability. Arrows in the group-wise reference indicate regions where significant differences can be observed between group-wise and majority vote reference choices.

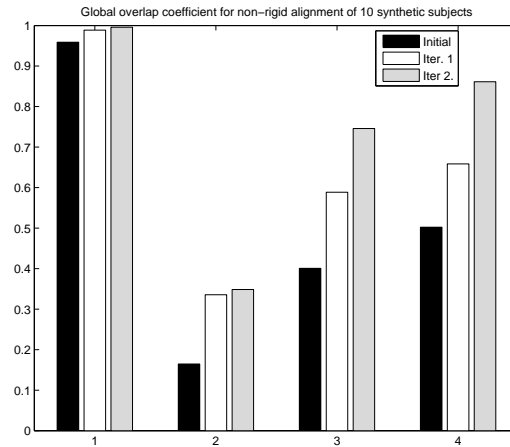


Figure 9: Evolution of the overlap coefficient defined in Equation 3.5 over consecutive alignment steps for each label (1 = background, 2 = Cerebro-spinal Fluid, 3 = Gray Matter, 4 = White Matter). The population considered is a synthetic population of 10 subjects generated using random B-Spline transformations. Two iterations of the non-rigid registration algorithm described in Section 3.1.2. were performed. Each iteration includes the estimation of the probabilistic reference and the dense non-rigid alignment of each subject on this reference.

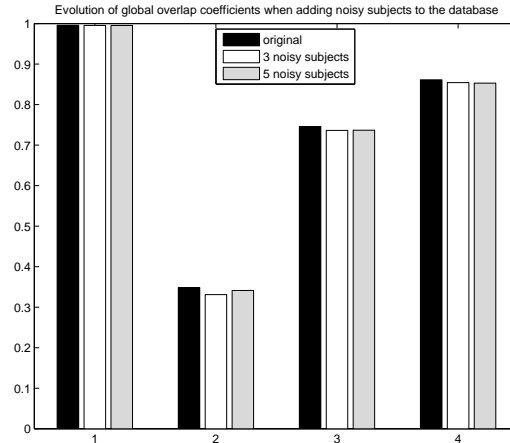


Figure 10: Evolution of the overlap coefficient defined in Equation 3.5 for each label (1 = background, 2 = Cerebro-spinal Fluid, 3 = Gray Matter, 4 = White Matter) when noisy subjects are added to the database. Outliers are produced using a random noise on each voxel (all voxels are mapped to an arbitrary label between 1 and 4 with a probability of 0.2).